

Memetic algorithms for ligand expulsion from protein cavities

J. Rydzewski¹ and W. Nowak^{1, a)}

*Institute of Physics, Faculty of Physics, Astronomy and Informatics,
Nicolaus Copernicus University, Grudziadzka 5, 87-100 Torun,
Poland*

(Dated: 2 July 2015)

Ligand diffusion through a protein interior is a fundamental process governing biological signaling and enzymatic catalysis. A complex topology of channels in proteins leads often to difficulties in modeling ligand escape pathways by classical molecular dynamics simulations. In this paper two novel memetic methods for searching the exit paths and cavity space exploration are proposed: Memory Enhanced Random Acceleration (MERA) Molecular Dynamics and Immune Algorithm (IA). In MERA, a pheromone concept is introduced to optimize an expulsion force. In IA, hybrid learning protocols are exploited to predict ligand exit paths. They are tested on three protein channels with increasing complexity: M2 muscarinic GPCR receptor, enzyme nitrile hydratase and heme-protein cytochrome P450cam. In these cases, the memetic methods outperform Simulated Annealing and Random Acceleration Molecular Dynamics. The proposed algorithms are general and appropriate in all problems where an accelerated transport of an object through a network of channels is studied.

Keywords: diffusion paths, immune algorithms, random acceleration molecular dynamics, ligand-protein complex, rare events

^{a)}To whom the correspondence should be addressed: wiesiek@fizyka.umk.pl

I. INTRODUCTION

The process of a ligand recognition is one of the most critical steps in biological signaling². To pass a signal, the ligand usually binds to a specific receptor docking site which may be exposed or buried inside a receptor matrix. Ligand's residence time in the receptor is of crucial importance in regulatory processes⁶⁰. The entrance, docking and egress processes involve usually a complex migration through the channels or ligand accessible cavities. Properties of these pathways determine the efficiency rate of signaling. Uncovering the distributions of transport routes is important not only in understanding mechanisms of the signal transduction, but also in enzymatic catalysis, molecular diseases, evolution of bio-systems and drugs design³. The process of ligand dissociation may be studied computationally³⁹, but classical molecular dynamics (MD) suffers from a low probability of expulsions. To facilitate crossing of steric barriers by ligands and to increase the rare conformational events probability, numerous enhanced MD methods have been developed, i.e. Locally Enhanced Sampling (LES)^{11,40,42}, Targeted MD⁴⁹, Steered MD (SMD)²⁷, Temperature Accelerated MD⁵³, Supervised MD⁴⁸. For a review of the enhanced MD methods see, for example, Johnston *et al.*²¹.

The memetic algorithms proposed here - MERA and IA may be considered to be an extension of SMD with a dynamically adjusted direction of an expulsion force. Therefore, it is worth to mention that SMD in its original form is limited in acquiring ligand egress paths, because the expulsion force ($\vec{F}_{smd} = -\frac{1}{2}k\nabla[v t - (\vec{r} - \vec{r}_0) \cdot \vec{n}]$, where k is a spring constant, v is a constant pulling velocity, \vec{r} and \vec{r}_0 are current and initial positions of a pulled atom, and finally, \vec{n} is a pulling direction), acting on a ligand is kept constant during the whole simulation so the path is limited to a straight line coordinate. Such a coordinate must be assumed a priori and it is a drawback of SMD.

A remedy has been found by Lüdemann *et al.*³² who have introduced the Random Acceleration MD (RAMD) (formerly known as Random Expulsion MD). In RAMD the expulsion force direction is modified when the travelling ligand meets any steric obstacle^{64,65}, which made an enforced egress of a camphor from cytochrome P450cam possible with no prior knowledge of an exit pathway³². Despite its popularity^{23,26,30}, the RAMD algorithm has one drawback: a trade-off between artificial perturbation of the protein structure caused by ligand pulling and a low probability of a ligand expulsion. Usually, numerous trajectories

have to be computed to get reasonable statistics^{32,63}.

Here, two new algorithms suitable for studies of the ligand transport through complex protein channels are introduced: MERA and IA. Both find plausible expulsion paths from buried receptor docking sites in three test systems: a M2 muscarinic receptor channel (M2), a biotechnological enzyme nitrile hydratase (NHase) and a cytochrome P450cam (P450cam), which has been tested in the original RAMD study by Lüdemann *et al.*³² (FIG. 1, 2).

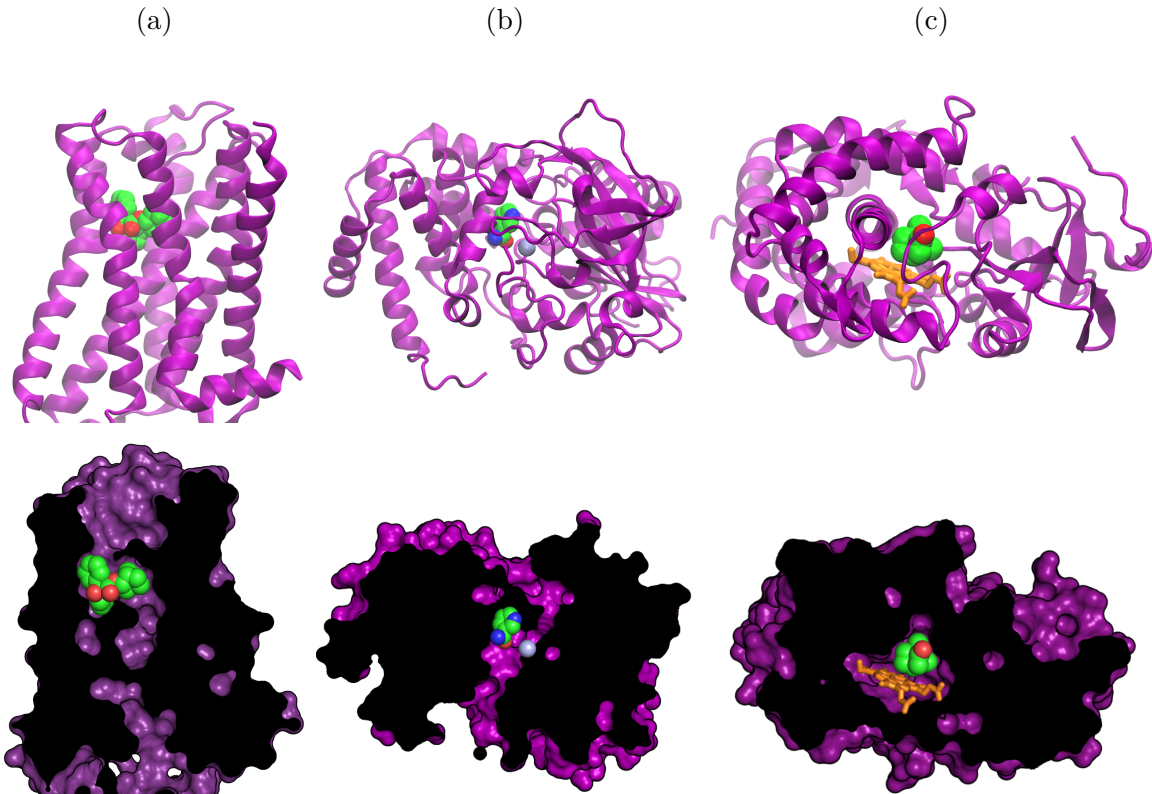


FIG. 1: The structures of ligand-protein complexes: (a) M2-QNB, (b) NHase-NCA with cobalt (light steel blue) in the catalytic center, (c) P450cam-CAM with heme (orange).

This article is organized as follows. The **Methods** section describes methods belonging to two different categories: **free energy independent** - RAMD, MERE and **free energy dependent** - IA and its variants IA-SW, IA-RMHC. Next, in the **Test systems and modeling protocols** section details about the simulation and modeling parameters of M2-QNB, NHase-NCA, P450cam-CAM are given. Parameters for RAMD, MERE and IA are presented in the same section. In the **Results and discussion** section results of egress pathways for three representative model systems of increasing complexity of the plausible

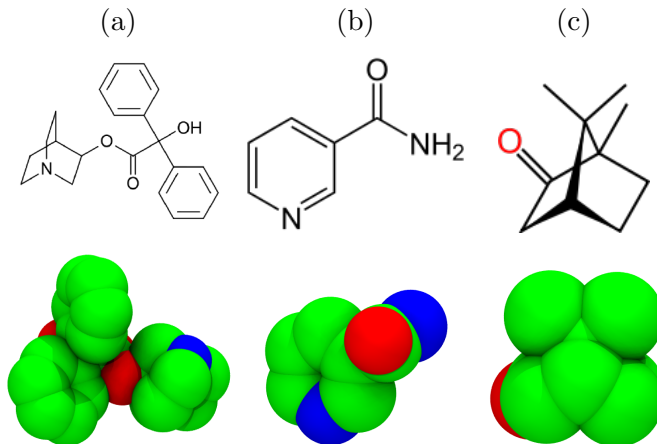


FIG. 2: The structures of ligands: (a) 3-Quinuclidinyl benzilate (QNB), (b) nicotinamide (NCA), (c) camphor (CAM).

exit channel, acquired from the studied algorithms, are shown and discussed. The MERE, IA, IA-SW and IA-RMHC calculations are compared to the results gained from traditional algorithms: RAMD and Simulated Annealing (Appendix B: Simulated annealing). The **Conclusions** section briefly summarizes advantages of the memetic methods for studies of ligand dissociation from proteins and possible applications in chemical physics.

II. METHODS

A. Free energy independent methods

1. *Random Acceleration Molecular Dynamics*

RAMD is considered to be an extension of the SMD method. There are two main advantages of the RAMD algorithm: (i) it speeds up dissociation kinetics and (ii) allows to find various probable dissociation routes. Also no prior knowledge of an exit pathway is required. The RAMD algorithm follows a specific protocol:

1. A direction \hat{r} of the external force acting on the center of mass of the ligand is applied randomly in such manner that $\vec{f}_{ext} = f_0 \hat{r}$, where f_0 is a constant magnitude of the randomly chosen force.
2. The force is maintained for a predetermined number of steps, m . A ligand is expected

to move with a velocity exceeding the threshold velocity given by $v_t = r_t/m\Delta t$, where Δt is a time step and r_t is a specified minimum distance before the direction change. If this condition is not fulfilled, the external force direction is reassigned randomly.

Finding all the specific parameters (r_t , f_0 , m) for the ligand-protein complexes is not trivial. Vashisth and Abrams considered it to be a drawback of the RAMD method - since the adopted force f_0 should not be too high, a percentage of successful dissociation events was 19%-41%.⁶³

The original RAMD scheme has been implemented in this article, but with an additional constraint (3.) introduced to decrease a high number of the unsuccessful ligand's expulsions from a receptor:

3. The new random force direction is chosen when a distance travelled by the ligand during the current interval $d(m\Delta t_i)$ of a simulation is smaller than a distance traveled in the previous interval $d(m\Delta t_{i-1})$. This happens when some obstacle is on the ligand's egress route.

2. Memory Enhanced Random Acceleration Molecular Dynamics

MERA algorithm has been implemented by adding a non-markovian dependency to the RAMD scheme. The representation of the memory is equivalent to pheromones trails in the Ant Colony Optimization (ACO) algorithms⁹. The concept of ACO has been exploited recently in the computational chemistry context^{10,41}. In MERA, every ligand leaves the pheromone trail during simulation and that trail is used as information for the next ligands positions on the dissociation pathway. Thus, a single pheromone corresponds to the previous ligand position. In other words, the ligands initially travel randomly, and upon dissociating to a protein exterior they lay down pheromone trails. If further ligands find this path, they are keen not to keep travelling at random, but to follow the trail, which leads outside. Because of that, a ligand must experience not only a stochastic force in a random direction $f_0\hat{r}$, but also a force directed to the most dense concentration ρ of pheromones $f_1\hat{k}$.

The initial distribution of pheromones in a protein is a key element of the MERA method and therefore has a big impact on calculations. A reasonably good guess is obtained by running exploratory LES¹¹ simulation or by collecting previous results of RAMD simulations.

During MERA simulations some paths may exhibit steric clashes and cavity traps. To eliminate these rare pathways pheromone refreshing is applied. Even if MERA produces such an expulsion, refreshing will dissolve those pheromones in order to eliminate a high driving force and not confuse consecutive ligands. The refreshing takes two steps to be completed:

1. In order to eliminate rarely visited paths, the pheromone trails continuously evaporate. All the pheromone concentrations are reduced according to geometrical scheme $\rho_i = \rho_i(1 - q)$, where ρ_i is concentration of the i -th pheromone and q is a damping factor (set to 0.01).
2. The concentrations of the pheromones are averaged within groups. The Shepard approximation algorithm⁵¹ is used with a KD-trees structure for finding the nearest neighbors of interpolated pheromone points. This optimization reduces the complexity of the Shepard approximation method to $O(N \log N)$, where N is a number of interpolated point's neighbors¹². Instead of the default interpolation kernel, the so-called Liszka's kernel has been considered³¹ (Appendix A: Shepard approximation kernels).

B. Free energy dependent methods

Computational physics problems can be expressed in terms of optimization techniques. In such cases, solving an optimization problem is equivalent to finding an extreme argument of a multivariate fitness function. In practice, the approximate results are sought, because optimization search techniques rarely converge to the global solution. Finding egress pathways may be formulated in terms of optimization problems. In the following section an optimization technique, IA, and its variants IA-RMHC and IA-SW, using two different local learning methods are introduced.

1. *Empirical interaction free energy function*

A fitness evaluation is the most crucial part of scoring-based metaheuristics i.e. evolutionary algorithms^{13,36} and simulated annealing²⁵. We used the scoring function that is non-bonded interaction energy between a ligand and a protein. For an egress process, the lower interaction energy between the ligand and the protein, the better solution. In

our implementation as the scoring function we considered the Hammett linear free energy relation^{17,18}. Moreover, the desolvation energy is computed using the partial volumes of atoms V scaled by the atomic solvation parameter S .

The interaction energy ΔG consists of four terms:

$$\begin{aligned}\Delta G = & \Delta G_{vdw} \sum_{i < j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) \\ & + \Delta G_h \sum_{i < j} \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right) \\ & + \Delta G_{elec} \sum_{i < j} \frac{q_i q_j}{\epsilon(r_{ij}) r_{ij}} \\ & + \Delta G_{sol} \sum_{i < j} (S_i V_j + S_j V_i) \exp(-r_{ij}^2 / 2\sigma^2)\end{aligned}\tag{1}$$

where the four $\Delta G_{(.)}$ coefficients on the right-hand side are empirically determined using the linear regression analysis from a set of ligand-protein complexes. These empirical data were taken from^{14,38}. The summations are performed over all ligand indices i and all protein indices j (in other words, for all pairs of ligand-protein atoms) in order to avoid energy doubling.

The first interaction energy term is the Lennard-Jones 12-6 potential. The second term is hydrogen bond energy and it is modeled by the Lennard-Jones 12-10 potential²². A , B , C and D are the matrices calculated to mimic the depth of the Lennard-Jones potential well and the equilibrium bond distance for homogeneous pair of atoms. The third term describes the Coulombic potential, where q is the charge of a given atom. The distance-dependent dielectric constant variable is modeled by the sigmoidal function of Mehler and Solmejer^{34,55,56}

$$\epsilon(r) = a + \frac{b}{1 + k \exp(-\lambda br)}\tag{2}$$

where $b = \epsilon_0 - a$ and $\epsilon_0 = 78.4$ (dielectric constant in water, in 25°C), $a = -8.5525$, $k = 7.7839$ and $\lambda = 0.003627 \text{\AA}^{-1}$. The last term in equation (1) represents the desolvation energy. In this work we used the partial atomic volume data from the paper of Stouten *et al.*⁵⁷. The evaluation of desolvation energy was the following: for every ligand's atom, partial atomic volumes of protein's atoms were scaled by the exponential function and then

summed. In this way we obtained the percentage of the protein’s surface, which surrounds a ligand. Such a result was then multiplied by an atomic solvation coefficient of ligand’s particular atom, which finally yields the desolvation energy. In other words, the last term describes to what extent the protein buries the ligand in its interior⁶⁶.

2. Immune algorithms

An immune algorithm is a computationally intelligent system inspired by the principles of the vertebrate immune system. Implementation of our IA algorithm belongs to a category of Clonal Selection Algorithms (CSA). CSA are population-based algorithms, which need two core components: cloning and hypermutation operators. The first triggers the growth of a new population, whereas the last can be seen as an optimization of the local search procedure that leads to a faster maturation during the learning phase⁴³. The applications of the IA algorithms have been recently exploited in various fields of computational chemistry and biophysics, i.e. protein structure prediction^{1,6,7}. IA involves an iterative process of mimicking the evolution of individuals belonging to the population. The scheme of the evolution is the following: (i) initially, the population is selected randomly, (ii) then the adaptation of each individual is calculated based on the scoring function, (iii) in the selection step individuals (i.e. the best in term of lowest energy score) are chosen from the population, (iv) selected individuals undergo evolutionary operations (i.e. mating, cloning, mutation, hypermutation), which reproduce a new population. The described process is repeated until the stop criterion is fulfilled.

IA works on the coded version of problems. Mapping similar to Hart *et al.*¹⁹ was used. The chromosome in our memetic algorithms (Scheme 1) consists of seven real numbers genes, which fully describe the translation (x, y, z) and orientation ($\hat{x}, \hat{y}, \hat{z}, \hat{w}$) of a ligand in the three-dimensional space.

Scheme 1: The representation of ligand in chromosome

x	y	z	\hat{x}	\hat{y}	\hat{z}	\hat{w}
-----	-----	-----	-----------	-----------	-----------	-----------

The selection mechanism depends on fitness of individuals. The better the fitness of a particular individual, the higher the probability of reproducing in a next generation. Selec-

tion strategies are based on Darwin’s law of natural selection⁸, due to which individual with better fitness survives. The roulette selection scheme was used:

$$p_i = \frac{f_i}{\sum_{j=1}^n f_j} \quad (3)$$

where f is the fitness of i -th individual and p is the probability of reproducing the i -th individual. When the selection is over, a crossover and then a mutation are performed on random members of a population, according to user-defined probabilities. The two-point crossover is used, with breaks occurring only between genes, never within. Individuals created during the crossover replace their parents in the population to keep its size constant. The mutation is performed by adding a real number (taken from the Cauchy distribution) to a randomly chosen gene in the chromosome.

The most substantial aspect of the implemented IAs is the usage of the restarts. The maturation process is executed by killing all individuals but one, in the population after several epochs. Then, the best individual from the killed population is cloned into the next stage of evolution. The current population consists of slightly disturbed, the best-adapted individuals from the previous evolution. Based on these principles, the cloning operator in the IA algorithm was implemented.

3. *Lamarckian adaptation*

Described IA falls into the category of global optimization techniques, but in some cases it is required to use both levels of optimization: local and global. The application of a local search metaheuristics provides an effective and robust method of hybrid optimization. Frequently, the global search does not sample sufficiently well the search space, which results in approximate solutions. The hybrid algorithms allow the separation of global and local searching, which in many cases helps to adjust an algorithm to a problem. In addition, it enables the rapid exchange of local search techniques: the global search is performed by default, but the local search technique can be chosen depending on the problem. Such hybrid techniques are often called Lamarckian and it is allusion to Jean Baptiste de Lamarck’s assertion that phenotypic characteristics acquired during an individual’s life can become heritable traits²⁹.

4. *Local search operator: hypermutation*

The local search is probably the oldest search metaheuristics⁵⁹. It starts with the initial solution, usually obtained with the global search. In each iteration the current solution is replaced by a better one (in terms of the fitness score).

We based our hybrid immunological methods (IA-RMHC, IA-SW) on two stochastic local searches: the Solis-Wets method (SW)⁵⁴ and Random Mutation Hill Climbing (RMHC)^{37,52}. The RMHC scheme is rather simple: the current individual is replaced only if a randomly generated neighbor is better adapted. The algorithm proposed by Forest and Mitchell has been used³⁷. In the case of ligand-protein complexes, a random neighbor is created by the Cauchy mutation of the current individual. This allows to observe an analogy between local search and the Lamarckian evolutionary hypothesis. The individual which undergoes the Cauchy’s mutation, actually adapts environmentally. These adaptations are passed to the next generation by the crossover (the mating scheme).

By default, the lamarckian adaptation is applied to each individual in a population. Such a decision is understandable, when calculation costs are not too high. Furthermore, the use of the local search for all of individuals, may cause the so-called over-fitting. We used the simplest selection scheme of the individuals. To adapt they were randomly chosen with defined probability (0.67).

SW is a more complex algorithm than RMHC. In SW, the sampling domain is dynamically adjusted to a success rate of finding the better solution⁵⁴. Only a few scientific studies attested that SW can be used as a local search technique^{14,19,38}. We have used local search (RMHC, SW) as a method of hypermutation, which means that such a step provides faster adaptation during the learning phase.

III. TEST SYSTEMS AND MODELING PROTOCOLS

5. *Structure of the program*

The main advantage of our program is a capability to compute ligand egress paths during a MD simulation. We used NAMD 2.9 code⁴⁶ to perform MD simulation with CHARMM 2.7 force field⁴. The communication between NAMD and the implemented program is done via NAMD’s feature, External Program Forces, which is an interface to calculate the external

forces. The implementation of all the methods presented in this study was performed in the C++ programming language⁵⁸ (standard 11) with boost library²⁴. As a random number generator, we used Mersenne twister of M. Matsumoto and T. Nishimura³³.

6. *Simulation parameters*

In all MD simulations a cutoff of 12Å for non-bonded interactions was applied and a 1 fs step size was used. Periodic boundary conditions were applied and full-system particle mesh Ewald periodic electrostatics was computed every time step. Langevin dynamics with a damping coefficient of 5 ps⁻¹ and Langevin piston algorithm (the piston period and decay were set to 200 fs and 50 fs, respectively) were used to keep temperature at 300 K and pressure at 1 atm.

7. *Muscarinic receptor M2*

The first tested ligand-protein complex was the M2 muscarinic receptor^{16,28}. M2 is a metabotropic receptor and its main agonist is an acetylcholine. The channel inside the M2 receptor is 33Å long. In the studied structure (PDB ID: 3OUN) the conformation of the M2 receptor remained inactive. Inside the M2 receptor resides a bulky (nearly 60 atoms) QNB ligand, which has a dissociation time longer than that of a native agonist - acetylcholine^{16,28}. The QNB ligand in this study was parameterized using the ParamChem server^{61,62} that helps in generating the force field parameters of small organic molecules. M2 was embedded in 1-Palmitoyl-2-oleoyl-*sn*-glycero-3-phosphoethanolamine (POPE) membrane using the VMD package²⁰. The M2 structure in membrane was solvated by the TIP3P water model (size of the whole box: 93.20Å×92.74Å×107.46Å). An electrically neutral environment was achieved by adding 0.15 mol/L potassium chloride ions.

Firstly, 0.5 ns equilibration of the POPE membrane was performed with 1 fs step size. During POPE equilibration velocities were reassigned every 1000 steps. Then, the whole system was minimized (1000 steps) and equilibrated (0.5 ns) with the harmonically constrained M2 receptor. Finally, a 0.3 ns equilibration of the whole model was performed.

8. *Nitrile hydratase*

Nitrile hydratase (PDB ID: 1IRE) is a metallo-enzyme used in the industrial production of acrylamide and nicotinamide (NCA). The NCA Charmm27 parameters and topology files were taken from Peplowski *et al.*⁴⁴. The NHase-NCA complex was solvated in (73Å×81Å×74Å) TIP3P water box and electrically neutralized by adding 0.15 mol/L sodium chloride ions. The 1000 steps of minimization and 100 ps of the whole system equilibration were carried out before production runs of the expulsion simulations.

9. *Cytochrome P450cam*

Cytochrome P450 enzymes are heme-containing protein monooxygenases that are related to the synthesis and degradation of many physiologically crucial compounds and in the degradation of xenobiotics¹⁵. We studied bacterial cytochrome P450cam from *Pseudomonas putida* docked with heme and its natural substrate - camphor (PDB ID: 2CPP). The active site is completely buried in the protein interior, thus it is necessary for the cytochrome P450cam to undergo structural changes in order to allow a substrate access and a product exit. The standard Charmm27 force field was used and CAM electric charges were collected from Schöneboom⁵⁰. The whole P450cam was solvated with a (77Å×75Å×57Å) TIP3P water box and electrically neutralized by adding 0.15 mol/L sodium chloride ions. Likewise, 1000 steps of minimization and 100 ps of the whole system equilibration were performed.

10. *Memetic algorithms parameters*

In all the test systems studied with IA, IA-SW and IA-RMHC, the size of all populations were set to 20. Mutations were applied to a randomly chosen individual from the population with the probability of 0.02. The mean and spread of Cauchy mutations were set to 0 and 1, respectively. The process of adapting the whole population was stopped after 20 epochs. The mating rate was set to 0.8 and the reproduced offspring were replacing the parents in the population. The hypermutations were applied to 0.67 of the whole population with the probability of 1 during 10 iterations. The sampling horizon from which all individuals were sampled was 8Å (M2-QNB), 15Å (NHase-NCA) and 10Å (P450cam-CAM).

In SA the temperature was reduced geometrically in cooling with damping parameter set to 0.7. The initial value of temperature-like control parameter in cooling was set to 50.

The LES simulation (500 ps long) in the initial distribution of pheromones in MERA-LES with 15 copies of corresponding ligands. The temperature and mass of these copies during simulation were reduced. In MERA MD, the initial distribution of pheromones was generated by prior 10 RAMD simulations with parameters set identically to latter MERA MD (see TABLE I-III). All the parameters regarding the dragging force constant f_0 , the frequency of direction choosing m and the number of dissociation simulations N are enlisted in TABLE I-III.

IV. RESULTS AND DISCUSSION

We have studied ligands' enforced diffusion paths in 3 model systems: GPCR M2 muscarinic receptor, enzyme nitrile hydratase, cytochrome P450cam. In these proteins the channels accessible to natural ligands show increasing complexity (FIG. 1). We have used the new algorithms proposed here: MERA, IA, IA-RMHC, IA-SW and, for the reference, the RAMD method and Simulated Annealing. The egress paths of the M2 antagonist QNB, NHase product nicotinamide and P450cam substrate camphor were calculated. At the beginning, all the ligands were placed inside the proteins docking sites. The resulting egress paths of successful ligands' exits are schematically presented in FIG. 3. The times of the longest trial egress simulations were 2.94 ps, 6.15 ps and 12.5 ps for M2-QNB, NHase-NCA and P450cam-CAM, respectively.

In order to assess the efficiency of methods, a simple statistics was collected for each model system studied. The success rate (%) defined as a ratio of a number of "successful" exits of the ligand from the protein docking site to a number of all computational trials undertaken has been estimated^{5,63}. We have monitored travel time (t) until a ligand reached the exterior. Expulsion time of RAMD simulations is short (TABLES I-III) due to the additional constraint applied to RAMD: the new random force direction is chosen, when the distance travelled by the ligand in the current interval $d(m\Delta t_i)$ is smaller than the distance travelled in the previous interval $d(m\Delta t_{i-1})$. The distance travelled by a ligand (d) was calculated by summing all partial distances which the ligands' center of the mass travelled during the simulation procedure.

Work done by a force used to expel a ligand from the protein interior (\vec{F}_{smd}) is a useful measure of obstacles met during the ligands' voyage. This work has been calculated as

$$W = \int \vec{F}_{smd} d\vec{s} \quad (4)$$

and collected for an each pathway. Along averaged data for these characteristics, their minimum and maximum values were recorded too. Data for M2-QNB, NHase-NCA and P450cam-CAM are presented in TABLE I, TABLE II and TABLE III, respectively.

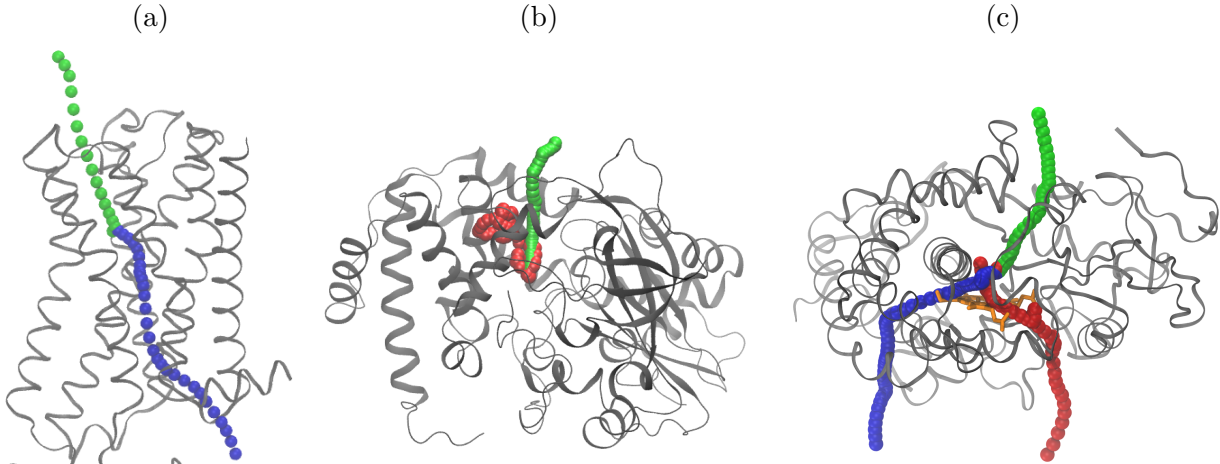


FIG. 3: The ligands' egress pathways from the docking sites of the receptors. (a) M2-QNB complex: pathways PW1 (green) and PW2 (blue), (b) NHase-NCA complex: pathway PW (green), example of an error (red), (c) Cytochrome P450cam-CAM complex: pathways PW1 (blue), PW2 (green), PWC (red).

11. *GPCR muscarinic receptor M2*

In M2-QNB, an antagonist is buried some 20Å inside a simple, regular cylindrical cavity (see Figure 1a). The MERA-LES algorithm based on LES pre-calculated positions of the ligand, gave the best success rate, 100%. When RAMD pre-calculations are used to set the initial pheromone distribution, instead of LES (MERA MD), the success rate drops to 66%. The success rates for the immune algorithms were also high (60%, 90% and 80% for IA, IA-RMHC and IA-SW respectively). SA was less successful in finding plausible paths of QNB in the M2 receptor: for PW1, the success rates were 44% and 77%, respectively. For

TABLE I: Characteristics of the expulsion pathways in the M2-QNB complex. The results of simulations for the dragging force constant - f_0 [kcal/molÅ], m - the frequency of choosing the next direction, N - the number of simulations, N_{pw} - the number of successful expulsions which occur on a corresponding pathway and % - efficiency rate. The minimal (Min.), maximal (Max.) and average (Avg.) values for time of expulsion t [ps], work w [kcal/mol] and distance travelled d [Å] are shown.

Algorithm	Pathway	f_0	m	N	N_{pw}	%	Min. t	Avg. t	Max. t	Min. d	Avg. d	Max. d	Min. w	Avg. w	Max. w
RAMD	PW1	5	10	20	9	45	0.86	1.11	1.35	26.56	30.15	37.30	75.6	100.4	126.3
RAMD	PW2	5	10	20	3	15	0.77	1.57	2.34	15.92	43.86	64.14	78.3	118.7	179.9
RAMD	PW1	10	10	20	5	25	0.45	0.72	0.99	24.84	38.39	65.73	193.0	260.4	408.5
RAMD	PW2	10	10	20	6	30	0.98	1.24	1.44	43.84	60.57	79.80	326.2	442.6	595.9
MERA MD	PW1	5	10	9	6	67	0.75	1.11	1.65	21.15	29.16	36.59	71.7	93.9	115.6
MERA-LES	PW1	5	10	10	10	100	0.64	0.9	1.14	18.18	25.49	32.77	63.7	89.7	118.7
SA	PW1	5	10	9	4	44	0.65	0.94	1.37	17.29	26.82	42.58	61.2	91.3	143.4
SA	PW2	5	10	9	2	22	1.28	1.36	1.49	34.25	40.15	46.04	118.7	138.0	157.2
SA	PW1	5	30	9	7	77	0.65	1.02	1.51	18.80	30.06	41.79	60.9	116.1	153.1
SA	PW2	5	30	9	1	11	0.99	0.99	0.99	27.93	27.93	27.93	114.4	114.4	114.4
IA	PW1	5	50	20	12	60	0.55	0.71	0.97	18.26	24.47	31.94	70.8	111.4	148.5
IA-RMHC	PW1	5	50	10	9	90	0.51	0.79	1.18	17.39	26.58	39.08	83.4	117.6	182.9
IA-SW	PW1	5	30	10	8	80	0.57	0.85	1.25	17.76	28.57	41.02	80.0	126.1	164.4

the pathway PW1, RAMD yields the success rate of 45%, which drops down to 25% when the dragging force constant is increased (i.e. when $f = 10$ kcal/molÅ is used instead of an optimum $f = 5$ kcal/molÅ, TABLE I).

The shapes of the PW1 egress paths (FIG. 3a) are similar in all the algorithms studied. We claim that it is a natural way for QNB to reach an exterior of M2. However, the average length of travel, as indicated by the parameter Avg. d in TABLE I, is the shortest for the path predicted by MERA MD and IA (25Å). For RAMD the ligand had to travel approximately 30Å before reaching the receptor's exterior region. One should note that for QNB, RAMD and SA predict the alternative exit path PW2 - this is, in our opinion, a highly unrealistic route leading to the interior of the M2 receptor embedded in the membrane. We are not aware of any experimental data that might support such an option. The MERAs and IAs methods predict only one type (PW1) of the exit path for the QNB ligand, which is perhaps the only correct option.

In the recent paper by Kruse *et al.*²⁸, authors have studied QNB diffusion paths in the M3 muscarinic receptor which has a similar GPCR structure. Numerous classical very long MD simulations (nearly 25 microseconds) were used to estimate pathways for the spontaneous QNB association with the M3 receptor. The paths found in that paper are qualitatively similar to the PW1 path, which supports utility of MERAs and IAs in studies of ligand and GPCR interactions.

A comparison of average work done by the expulsion force in all algorithms tested is also in favor of MERAs and IAs. The work performed by the expulsion forces is the lowest in these cases, and has a value of 90 kcal/mol, while the best value for Avg. w in RAMD is 100 kcal/mol. Thus, we expect that the perturbations of a receptor structure induced by the process of an enforced ligand dissociation is smaller in MERAs and IAs than in RAMD, which is another advantage of the proposed memetic algorithms (FIG. 4). It is worth noting that Avg. w for the unrealistic PW2 paths calculated by RAMD may be as high as 442 kcal/mol (TABLE I).

12. *Enzyme nitrile hydratase*

All the algorithms studied here were applied to calculate exit paths for nicotinamide (also referred to as vitamin PP) docked inside the NHase spherical pocket, near the catalytic center of this metaloenzyme. NCA is buried approximately 40Å beneath the protein exterior surface. The channel is not long, but in contrast to the M2 receptor, it is bent and more sophisticated, thus the NCA diffusion problem is more challenging. All methods led to the same exit pathway, PW (FIG. 3b). It is in a good agreement with our previous LES based calculations with other amides⁴⁵. The results on the successful rate and calculated paths characteristics are presented in TABLE II.

MERA MD shows the moderate successful rate in this system, 44%. However, when information on possible ligand's locations, obtained from pre-calculated LES is added, the success rate is 90% (TABLE II). Such an increase in the percentage of successful trajectories may indicate, that the LES algorithm, due to lowered potential energy barriers, performs the initial sampling of NHases interior better than the stochastic walk RAMD. The IAs algorithms show an excellent success rate, despite the lack of pre-calculated hints: IA - 90% , IA-RMHC - 100% and IA-SW - 100%. There is no difference in the success rates of

IA-RMHC and IA-SW, because the docking pocket in NHase is a wide, spherical cavity and the ligand which penetrates NHase interior has a lot of free space to change its position. In this case, correction of the radius of the sampling domain in the SW local search is not necessary. Therefore, SW behaves similarly to the RMHC hypermutation. The success rate of SA is low: 27% and even worse is RAMD: 0%-14%. If the expulsion path is found, the mean travel time Avg. t in RAMD is 5 times higher than that of MERAs (5.3 ps and 1.2 ps, TABLE II). It shows clearly that the algorithms with a more advanced sampling scheme find an optimal, in terms of receptor's geometry, egress pathway. The random walk RAMD method and SA do not find optimal pathways and, more importantly, the search for any possible dissociation pathway is complicated.

The distances d calculated by all the memetic methods are similar (TABLE II) and vary from 43Å (MERA-LES) to 77Å (MERA MD). The longest paths were predicted using RAMD - it took a 177Å journey before the exit has been found.

Since the expulsion force value was the same in all methods used, the work done by this force to dissociate the ligand depends on the travelled path length. The lowest value of 69 kcal/mol was found for IA-SW and the largest value of 482 kcal/mol resulted from RAMD calculations.

TABLE II: Characteristics of the expulsion pathways in the NHase-NCA complex. The results of simulations for the dragging force constant - f_0 [kcal/molÅ], m - the frequency of choosing the next direction, N - the number of simulations, N_{pw} - the number of successful expulsions which occur on a corresponding pathway and % - efficiency rate. The minimal (Min.), maximal (Max.) and average (Avg.) values for time of expulsion t [ps], work w [kcal/mol] and distance travelled d [Å] are shown.

Algorithm	Pathway	f_0	m	N	N_{pw}	%	Min. t	Avg. t	Max. t	Min. d	Avg. d	Max. d	Min. w	Avg. w	Max. w
RAMD	PW	10	50	14	2	14.2	5.35	5.75	6.15	49.72	117.48	185.24	247.9	482.0	716.1
RAMD	PW	15	30	7	0	0	-	-	-	-	-	-	-	-	-
MERA MD	PW	10	100	9	4	44.4	1.15	3.71	6.00	38.96	77.64	153.78	121.9	242.1	473.3
MERA-LES	PW	10	100	10	9	90	1.15	1.61	1.95	33.85	42.66	55.03	82.1	106.5	134.7
SA	PW	10	100	11	3	27	1.50	2.88	5.3	52.43	56.45	63.88	178.1	193.0	200.6
IA	PW	10	50	10	9	90	1.30	1.68	1.90	31.45	46.83	54.41	49.5	76.5	95.6
IA-RMHC	PW	10	50	15	15	100	1.10	1.66	1.95	21.07	44.49	54.56	36.7	72.5	89.0
IA-SW	PW	10	50	10	10	100	1.55	1.63	1.80	32.57	43.33	49.18	44.0	69.5	77.4

13. Cytochrome P450

As the last and the most complex test case, cytochrome P450cam from *Pseudomonas Putida* has been used (FIG. 1c) since the same enzyme was studied in the article where RAMD has been introduced by Lüdemann *et al.* P450cam is a globular heme-protein engaged in the synthesis and degradation of numerous metabolites and xenobiotics. Camphor is one of its natural substrates. During the P450cam activity CAM enters a distal pocket buried inside the enzyme and located close to the heme moiety. Lüdemann *et al.*³² used RAMD to study all possible CAM dissociation paths starting from this initial position. After running hundreds of simulations the authors have found three distinct routes (1, 2, 3), but for the path no. 2 three variants were discriminated. The results of our RAMD calculations for P450cam basically agree with those obtained by Lüdemann *et al.*: we have found three groups of pathways as well. Our PW1 path (FIG. 3c) is identical like that presented in Lüdemann *et al.* The same refers to our PW2 and the pathway no. 2a. However, we were not able to observe the pathway PW3 as reported earlier, perhaps due to rather limited statistics. Instead we have found the third pathway in P450cam, which corresponds to so-called water channel (PWC, FIG. 3c). PWC has been already postulated by Poulos in the article on the P450cam structure⁴⁷. One should note that the absence of PWC in the RAMD study and the rare appearance of this variant in our RAMD calculations are clearly related to the nature of the RAMD algorithm: this channel is not easily accessible, since PWC is protected by the rigid and bulky heme group and an expulsion force direction selection scheme in RAMD is too simplistic to overcome this obstacle. Our more advanced algorithms perform better in this case.

The detailed results on assessment of all algorithms in the case of the cytochrome P450cam are presented in TABLE III. One can see that 60% of the RAMD trials have led to the CAM ligand exit. This result has been achieved thanks to application of a rather big dragging force constant ($f = 10$ kcal/molÅ, TABLE III). If a dragging force constant of $f = 5$ kcal/molÅ was used in RAMD, the success exit trajectories were more rare (20%). However, in case of the memetic algorithms postulated here, the success rates were much better. MERA MD gave the 80% success rate and using additional pre-calculated data from LES (MERA-LES) has provided 100% success trajectories with the CAM ligand exits (PW1 30%, PW2 70%). All the variants of IA-based algorithms (IA, IA-RMHC and IA-SW) gave

a collective 100% success rate (TABLE III) as well. Except for the IA-SW algorithm, all the methods predict that the PW2 path dominates in the CAM transport in cytochrome P450cam, in accordance with the earlier report³². The success rate of SA for this model system was 60% (three paths were predicted, including PWC).

The analysis of time of ligands' travel within P450cam (Avg. t , TABLE III) is also informative. Time of 7.30 ps, $f = 5$ kcal/molÅ for PW2 given by RAMD is in contrast to time of 1.15 ps observed also by RAMD on PW2 (TABLE III). This shortening of the simulation has been achieved by using a higher dragging force constant $f = 10$ kcal/molÅ. However, one should always check to what extent the protein interior is deformed by such a ligand expulsion, for instance by monitoring RMSD values and collisions of a ligand and amino acids belonging to a receptor (FIG. 4).

Our MERA MD algorithm needs 2.63 ps to find an exit for the same conditions (PW2, $f = 5$ kcal/molÅ), thus a total simulation time for MERA MD is approximately 3 times shorter than in RAMD simulations. A comparison of the distance travelled by the CAM ligand in both cases (Avg. d , TABLE III) favors MERA MD as well (43.8Å and 25.53Å for RAMD and MERA MD, respectively). Interestingly, the shortest distance path PW2 was predicted by the IA-SW method (15.44Å). This very short path has been quickly found (Avg. t , 1.53 ps) due to a variable sampling radius adopted in the IA-SW algorithm. This feature is particularly advantageous when the diffusion channel is very narrow, like in P450cam. All the other path calculation algorithms need numerous small trials before a successful force direction is gradually determined. Here, in IA-SW a larger sampling radius quickly brings information that larger area is available for a ligand. The work Avg. w (see Table 3) performed by the expulsion force along the PW2 path is the smallest one (66 kcal/mol). A relatively low work of 90 kcal/mol is needed to force the CAM ligand through PW2 as predicted by the MERA MD algorithm, but as much as 305 kcal/mol is required to run the ligand along the PW1 path by the high force RAMD.

Our analysis of the minimum, average and maximum work done during the enforced ligand egress in model protein systems has a deeper meaning than just checking a success rate of the memetic algorithms. One can infer from the data which pathway is the most optimal in sense of the interaction free energy and the least perturbing. Thus, the paths with the lowest work can be selected for further analysis of biological phenomena, such as the role of mutations in ligands diffusion dynamics or enzymatic selectivity improvement.

TABLE III: Characteristics of the expulsion pathways in the P450cam-CAM complex. The results of simulations for the dragging force constant - f_0 [kcal/molÅ], m - the frequency of choosing the next direction, N - the number of simulations, N_{pw} - the number of successful expulsions which occur on a corresponding pathway and % - efficiency rate. The minimal (Min.), maximal (Max.) and average (Avg.) values for time of expulsion t [ps], work w [kcal/mol] and distance travelled d [Å] are shown.

Algorithm	Pathway	f_0	m	N	N_{pw}	%	Min. t	Avg. t	Max. t	Min. d	Avg. d	Max. d	Min. w	Avg. w	Max. w
RAMD	PW1	10	100	10	2	20	1.30	2.03	2.75	33.06	49.83	66.60	287.6	417.6	270.4
RAMD	PW2	10	100	10	3	30	1.05	1.15	1.30	27.73	29.16	30.67	214.4	242.8	270.4
RAMD	PWC	10	100	10	1	10	2.05	2.05	2.05	39.96	39.96	39.96	305.2	305.2	305.2
RAMD	PW2	5	100	10	2	20	2.10	7.30	12.50	25.83	43.88	61.92	114.1	178.3	242.5
MERA MD	PW1	5	100	10	3	30	2.10	2.63	2.90	25.59	31.26	39.12	85.4	106.7	141.8
MERA MD	PW2	5	100	10	5	50	1.00	1.87	2.85	18.16	25.53	34.55	68.1	92.8	130.1
MERA-LES	PW1	5	100	10	3	30	2.50	3.88	5.35	29.20	38.15	49.86	90.2	115.6	150.6
MERA-LES	PW2	5	100	10	7	70	3.30	3.90	4.30	28.96	36.82	48.09	72.9	119.5	156.7
SA	PW1	5	100	10	1	10	4.00	4.00	4.00	57.47	57.47	57.47	218.5	218.5	218.5
SA	PW2	5	100	10	3	30	1.21	1.85	2.50	30.05	43.16	64.26	125.7	308.1	511.8
SA	PWC	5	100	10	2	20	1.45	2.88	4.30	35.85	52.83	69.80	286.8	300.9	315.1
IA	PW1	5	100	10	2	20	1.70	2.65	3.60	24.65	32.50	40.34	109.9	130.3	150.8
IA	PW2	5	100	10	8	80	1.00	1.61	2.60	20.59	24.25	36.89	92.8	109.8	152.6
IA-RMHC	PW1	5	100	10	1	10	3.10	3.10	3.10	47.95	47.95	47.95	187.9	187.9	187.9
IA-RMHC	PW2	5	100	10	1	60	1.45	2.90	4.30	22.00	43.84	55.51	97.5	174.9	215.6
IA-RMHC	PWC	5	100	10	3	30	3.45	4.68	5.30	50.73	66.43	77.64	190.9	254.7	296.8
IA-SW	PW1	5	100	10	1	10	2.10	2.10	2.10	25.72	25.72	25.72	117.1	117.1	117.1
IA-SW	PW2	5	100	10	3	30	1.20	1.53	1.7	8.95	15.44	20.98	32.0	66.2	90.8
IA-SW	PWC	5	100	10	6	60	2.20	2.89	3.40	21.53	24.18	27.64	92.1	99.3	116.1

The summary of the success rates for all algorithms is presented in TABLE IV. We have calculated a ratio of a number of successful paths, despite of their topology, to a number of all exit simulation attempts. The results show that the best average success rate (Avg. SR) of 96.7% is offered by the MERA-LES algorithm, but one need to remember that MERA-LES calculations were based on the pre-calculated pheromones distributions by LES. Since IAs do not require similar pre-calculations, they are the most successful in our study. The IA-RMHC and IA-SW methods display successful rate of 96.7% and 93.3%, respectively. The MERA MD algorithm has SR of 63.3% which is slightly better than SA (58.3%) and much better than RAMD (44.3%). Moreover, the new algorithms presented here show that

TABLE IV: Summary of the success rates SR [%] for all the ligand-protein complexes studied.

Algorithm	SR_{M2}	SR_{NHase}	$SR_{P450cam}$	Avg. SR
RAMD	60	14	60	44.6
MERA MD	66	44	80	63.3
MERA-LES	100	90	100	96.7
SA	88	27	60	58.3
IA	60	90	100	83.3
IA-RMHC	90	100	100	96.7
IA-SW	80	100	100	93.3

collision statistics and the RMSD values generated during 10 randomly chosen trajectories for each system tested, are much lower in IA, comparing to RAMD (FIG. 4). It is worth noting that despite the high percentage of successful dissociations, IA and MERA do not introduce excessive artifacts in the sampling of the protein conformations.

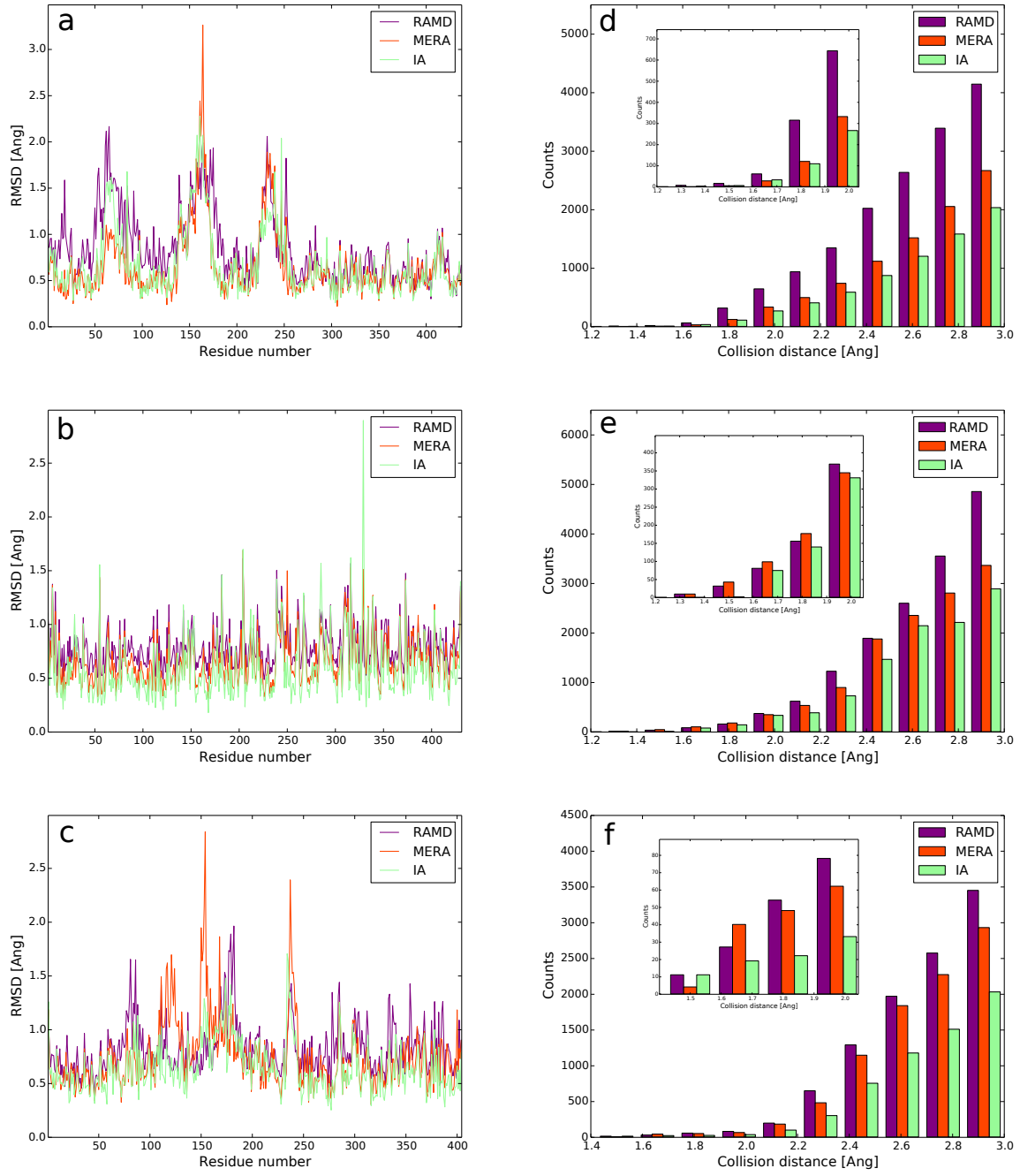


FIG. 4: RMSD per residue (a, b, c) and collision statistics (d, e, f) for M2-QNB (a, d), NHase-NCA (b, e) and P450cam-CAM (c, f) complexes from 10 randomly chosen trajectories for each RAMD, MERA and IA-RMHC expulsion simulations.

V. CONCLUSIONS

Molecular modeling in biology, biophysics or drug design often requires computation of small molecules' travel pathways within macromolecular matrices. Channels are exploited in the nature by ligands to reach docking or catalytic sites. Then ligands have to exit outside. The classical MD simulations contribute to prediction of such processes, however, simulations of diffusion may be extremely time-consuming. Some accelerated MD methods, such as RAMD, help to determine ligand pathways in an approximate, but more optimal way.

In this paper we have presented two types of new algorithms that alleviate these problems and improve ligand paths search substantially: MERA and IA. These memetic algorithms were tested on three protein-ligand systems showing the increasing complexity of the channels: a muscarinic receptor M2-QNB, an enzyme nitrile hydratase - nicotinamide, and cytochrome P450-camphor complex.

In the MERA algorithms a memory effect has been added to RAMD, since this idea was inspired by the ACO approach⁹. Calculations have shown that this method may help improve RAMD. Moreover, in our variant of MERA we have used the Locally Enhanced Sampling¹¹ to pre-calculate plausible positions of the ligand in a diffusion channel studied. The resulting MERA-LES method has shown a high success rate (96%). One should note that the LES pre-calculations require additional, but reasonable computing time.

The another group of new algorithms draws on the immune system. In the IA approach a target function based on ligand-host interaction energy is used to assess the direction of the expulsion force. In IA-SW a varying search radius is implemented. In the IA-RMHC a special iterative procedure of local search is used. The success rate of IA algorithms is also very good: 93-96% and they do not require any pre-calculations, as opposed to MERA. We recommend IA-SW for studies of diffusion in narrow channels.

The calculated distinct exit paths are reasonable in all systems: M2, NHase and P450cam. Their shapes are in agreement with the previous calculations. Since our paths are more smooth then the pathways calculated previously, we believe that they are more suitable for discussions of diffusion mechanisms and for planning new mutagenesis experiments. One may use snapshots from the calculated paths to optimize them further to Steepest Descent Paths.⁴⁰

We hope that the methods described here will enhance modeling of new, complex biophysical phenomena. The algorithms are general, easy to implement and not limited to protein studies. Both MERA and IA methods may be easily used, for example, in porous material science or ions transport through membrane ion channels.

ACKNOWLEDGMENTS

This research was supported in part (WN) by the NCN grant № N N202 262038. We thank for the computer time allocated by the Interdisciplinary Center for Modern Technologies, NCU. We would like to thank Rafal Jakubowski for useful discussions, Aleksander Balter and Magdalena Rydzewska for critically reading the manuscript.

Appendix A: Shepard approximation kernels

The Shepard approximation (SHA) is a deterministic method for a multivariate interpolation with a known scattered set of points. The values assigned to unknown points are calculated with a weighted average of values available at the known points⁵¹. The SHA formula for an interpolated value u for a certain argument \vec{x} is the following:

$$u(\vec{x}) = \frac{\sum_{k=0}^N w_k(\vec{x}) u_k}{\sum_{k=0}^N w_k(\vec{x})} \quad (\text{A1})$$

where a weight coefficient $w_k(\vec{x})$ called an interpolation kernel is given by the equation

$$w_k(\vec{x}) = 1/d(\vec{x}, \vec{x}_k)^p \quad (\text{A2})$$

and \vec{x} is an interpolated point, \vec{x}_k is scattered data; d defines a metric to calculate the distances, N stands for a number of scattered interpolating data and finally $p \in [0, 2]$ is a parameter. Although the equation (A2) is applicable for standard tasks, we have been unable to use it. Here, using approximation for known data which already have a pheromone concentration value leads to a singularity (because $d(\vec{x}, \vec{x}_k)$ is equal to 0). Instead of the default interpolation kernel, we considered so-called Liszka's kernel³¹

$$w_k^L(\vec{x}) = \sqrt{d(\vec{x}, \vec{x}_k)^2 + \epsilon^2} \quad (\text{A3})$$

where ϵ is an additional parameter.

Appendix B: Simulated annealing

SA is a stochastic metaheuristics of global optimization⁵⁹. It belongs to a wider class of adaptive random search methods¹⁹, because in order to create a new solution, the knowledge about past solutions is required. SA generates a new solution by stochastically deviating the current one. A deviation is accepted, if the Metropolis condition is fulfilled³⁵. In other words, the new solution is accepted if $f(x') - f(x'') < 0$ holds, where x' is the new solution and x'' is an old one. Otherwise, the solution is accepted according to the Boltzmann probability $P_B = \exp[(f(x'') - f(x'))/T]$, where T is a temperature or a controlling parameter. Typically T is reduced during annealing. Such a method of a decreasing temperature is often called a cooling scheme. The higher T value, the higher probability of accepting the worse (in the sense of scoring function) solution. During cooling, the probability of acceptance is reduced accordingly. Because of that the SA algorithm at the beginning performs a global optimization, and at the end rather small local changes.

The nature of finding a ligand's egress pathway from the receptor gives a perfect opportunity to use a restarting scheme in SA. The reason for that is an urge to perform SA in certain predetermined points of time. The direction of pulling adjusts to the geometry of a receptor-ligand complex and because of that, it is obvious that the SA optimization should be performed multiple times, one for each placement of a ligand in a receptor site. This methodology prevents from ending in a local optimum.

REFERENCES

- ¹Anile, A. M., Cutello, V., Narzisi, G., Nicosia, G., and Spinella, S., Nat Comput **6**, 55 (2007).
- ²Bhalla, U. S. and Iyengar, R., Science **283**, 381 (1999).
- ³Bhat, M., Biotechnol. Adv. **18**, 355 (2000).
- ⁴Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M., J. Comput. Chem. **4**, 187 (1983).
- ⁵Carlsson, P., Burendahl, S., and Nilsson, L., Biophys. J. **91**, 3151 (2006).
- ⁶Cutello, V., Morelli, G., Nicosia, G., Pavone, M., and Scollo, G., Nat Comput **10**, 91 (2011).

- ⁷Cutello, V., Nicosia, G., Pavone, M., and Timmis, J., IEEE T EVOLUT COMPUT **11**, 101 (2007).
- ⁸Darwin, C., *On the Origin of Species: By Means of Natural Selection* (The Floating Press, 2009).
- ⁹Dorigo, M. and Stützle, T., in *Handbook of Metaheuristics* (Springer, 2003) pp. 250–285.
- ¹⁰Dresselhaus, T., Yang, J., Kumbhar, S., and Waller, M. P., J. Chem. Theory Comput. **9**, 2137 (2013).
- ¹¹Elber, R. and Karplus, M., J. Am. Chem. Soc. **112**, 9161 (1990).
- ¹²Friedman, J. H., Bentley, J. L., and Finkel, R. A., ACM Transactions on Mathematical Software (TOMS) **3**, 209 (1977).
- ¹³Goldberg, D. E., *Genetic algorithms* (Pearson Education India, 2006).
- ¹⁴Goodsell, D. S. and Olson, A. J., Protein **8**, 195 (1990).
- ¹⁵Graham-Lorence, S., Peterson, J. A., Amarneh, B., Simpson, E. R., and White, R. E., Protein Sci. **4**, 1065 (1995).
- ¹⁶Haga, K., Kruse, A. C., Asada, H., Yurugi-Kobayashi, T., Shiroishi, M., Zhang, C., Weis, W. I., Okada, T., Kobilka, B. K., and Haga, T., Nature **482**, 547 (2012).
- ¹⁷Hammett, L. P., J. Am. Chem. Soc. **59**, 96 (1937).
- ¹⁸Hammett, L. P., Trans. Faraday Soc. **34**, 156 (1938).
- ¹⁹Hart, W. E., *Adaptive global optimization with local search*, Ph.D. thesis, University of California, San Diego (1994).
- ²⁰Humphrey, W., Dalke, A., and Schulten, K., J. Mol. Graphics **14**, 33 (1996).
- ²¹Johnston, J. M. and Filizola, M., in *G Protein-Coupled Receptors-Modeling and Simulation* (Springer, 2014) pp. 95–125.
- ²²Jones, J. E., Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character , 463 (1924).
- ²³Kalyaanamoorthy, S. and Chen, Y.-P. P., J. Chem. Inf. Model. **52**, 589 (2012).
- ²⁴Karlsson, B., *Beyond the C++ standard library: an introduction to boost* (Pearson Education, 2005).
- ²⁵Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P., Science **220**, 671 (1983).
- ²⁶Klvana, M., Pavlova, M., Koudelakova, T., Chaloupkova, R., Dvorak, P., Prokop, Z., Stsiapanava, A., Kutý, M., Kuta-Smatanova, I., Dohnalek, J., *et al.*, J. Mol. Biol. **392**, 1339 (2009).

- ²⁷Kosztin, D., Izrailev, S., and Schulten, K., *Biophys. J.* **76**, 188 (1999).
- ²⁸Kruse, A. C., Hu, J., Pan, A. C., Arlow, D. H., Rosenbaum, D. M., Rosemond, E., Green, H. F., Liu, T., Chae, P. S., and Dror, R. O. e. a., *Nature* **482**, 552 (2012).
- ²⁹Lamarck, J. B., *Zoological Philisophy* (Macmillan London, 1914).
- ³⁰Li, W., Shen, J., Liu, G., Tang, Y., and Hoshino, T., *Proteins* **79**, 271 (2011).
- ³¹Liszka, T., *Int. J. Numer. Meth. Eng.* **20**, 1599 (1984).
- ³²Lüdemann, S. K., Lounnas, V., and Wade, R. C., *J. Mol. Biol.* **303**, 797 (2000).
- ³³Matsumoto, M. and Nishimura, T., *ACM Transactions on Modeling and Computer Simulation (TOMACS)* **8**, 3 (1998).
- ³⁴Mehler, E. L. and Solmajer, T., *Protein Eng.* **4**, 903 (1991).
- ³⁵Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E., *J. Chem. Phys.* **21**, 1087 (1953).
- ³⁶Michalewicz, Z., *Genetic algorithms + data structures = evolution programs* (Springer, 1996).
- ³⁷Mitchell, M., Holland, J. H., and Forrest, S., in *NIPS* (1993) pp. 51–58.
- ³⁸Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., and Olson, A. J., *J. Comput. Chem.* **19**, 1639 (1998).
- ³⁹Nowak, W., in *Handbook of Computational Chemistry* (Springer, 2012) pp. 1127–1153.
- ⁴⁰Nowak, W., Czerminski, R., and Elber, R., *J. Am. Chem. Soc.* **113**, 5627 (1991).
- ⁴¹Oakley, M. T., Richardson, E. G., Carr, H., and Johnston, R. L., *IEEE/ACM Trans. Comput. Biology Bioinform.* **10**, 1548 (2013).
- ⁴²Orlowski, S. and Nowak, W., *J. Mol. Model.* **13**, 715 (2007).
- ⁴³Pavone, M., Narzisi, G., and Nicosia, G., *J. Global Optim.* **53**, 769 (2012).
- ⁴⁴Peplowski, L., Kubiak, K., and Nowak, W., *J. Mol. Model.* **13**, 725 (2007).
- ⁴⁵Peplowski, L., Kubiak, K., and Nowak, W., *Chem. Phys. Lett.* **467**, 144 (2008).
- ⁴⁶Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L., and Schulten, K., *J. Comput. Chem.* **26**, 1781 (2005).
- ⁴⁷Poulos, T. L., Finzel, B. C., and Howard, A. J., *Biochemistry* **25**, 5314 (1986).
- ⁴⁸Sabbadin, D. and Moro, S., *J. Chem. Inf. Model.* **54**, 372 (2014).
- ⁴⁹Schlitter, J., Engels, M., and Krüger, P., *J. Mol. Graphics* **12**, 84 (1994).
- ⁵⁰Schöneboom, J. C., Lin, H., Reuter, N., Thiel, W., Cohen, S., Ogliaro, F., and Shaik, S., *J. Am. Chem. Soc.* **124**, 8142 (2002).

- ⁵¹Shepard, D., in *Proceedings of the 1968 23rd ACM national conference* (ACM, 1968) pp. 517–524.
- ⁵²Skalak, D. B., in *ICML* (Citeseer, 1994) pp. 293–301.
- ⁵³So, M. R. and Voter, A. F., *J. Chem. Phys.* **112**, 9599 (2000).
- ⁵⁴Solis, F. J. and Wets, R. J.-B., *Math. Oper. Res.* **6**, 19 (1981).
- ⁵⁵Solmajer, T. and Mehler, E., *Int. J. Quantum Chem.* **44**, 291 (1992).
- ⁵⁶Solmajer, T. and Mehler, E. L., *Protein Eng.* **4**, 911 (1991).
- ⁵⁷Stouten, P. F., Frömmel, C., Nakamura, H., and Sander, C., *Mol. Simulat.* **10**, 97 (1993).
- ⁵⁸Stroustrup, B., *The C++ programming language* (Pearson Education India, 1995).
- ⁵⁹Talbi, E.-G., *Metaheuristics: from design to implementation*, Vol. 74 (John Wiley & Sons, 2009).
- ⁶⁰Tummino, P. J. and Copeland, R. A., *Biochemistry* **47**, 5481 (2008).
- ⁶¹Vanommeslaeghe, K. and MacKerell Jr, A. D., *J. Chem. Inf. Model.* **52**, 3144 (2012).
- ⁶²Vanommeslaeghe, K., Raman, E. P., and MacKerell Jr, A. D., *J. Chem. Inf. Model.* **52**, 3155 (2012).
- ⁶³Vashisth, H. and Abrams, C. F., *Biophys. J.* **95**, 4193 (2008).
- ⁶⁴Wang, T. and Duan, Y., *J. Am. Chem. Soc.* **129**, 6970 (2007).
- ⁶⁵Wang, T. and Duan, Y., *J. Mol. Biol.* **392**, 1102 (2009).
- ⁶⁶Wesson, L. and Eisenberg, D., *Protein Sci.* **1**, 227 (1992).